



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2017

Multi-word Adverbs – How well are they handled in Parsing and Machine Translation?

Volk, Martin ; Graën, Johannes

Abstract: Multi-word expressions are often considered problematic for parsing or other tasks in natural language processing. In this paper we investigate a specific type of multi-word expressions: binomial adverbs. These adverbs follow the pattern adverb + conjunction + adverb. We identify and evaluate binomial adverbs in English, German and Swedish. We compute their degree of idiomaticity with an ordering test and with a mutual information score. We show that these idiomaticity measures point us to a number of fixed multi-word expressions which are often mis-tagged and mis-parsed. Interestingly, a second evaluation shows that state-of-the-art machine translation handles them well – with some exceptions.

Posted at the Zurich Open Repository and Archive, University of Zurich
ZORA URL: <https://doi.org/10.5167/uzh-141846>
Conference or Workshop Item

Originally published at:

Volk, Martin; Graën, Johannes (2017). Multi-word Adverbs – How well are they handled in Parsing and Machine Translation? In: The 3rd Workshop on Multi-word Units in Machine Translation and Translation Technology (MUMTTT 2017), London, 14 November 2017, s.n..

Multi-word Adverbs – How well are they handled in Parsing and Machine Translation?

Martin Volk, Johannes Graën
University of Zurich
Institute of Computational Linguistics
volk|graen@cl.uzh.ch

Abstract

Multi-word expressions are often considered problematic for parsing or other tasks in natural language processing. In this paper we investigate a specific type of multi-word expressions: binomial adverbs. These adverbs follow the pattern *adverb + conjunction + adverb*. We identify and evaluate binomial adverbs in English, German and Swedish. We compute their degree of idiomaticity with an ordering test and with a mutual information score. We show that these idiomaticity measures point us to a number of fixed multi-word expressions which are often mis-tagged and mis-parsed. Interestingly, a second evaluation shows that state-of-the-art machine translation handles them well – with some exceptions.

1 Introduction

We work on the annotation of large corpora for linguistic research and information extraction. We noticed that multi-word adverbs often cause confusion to the PoS tagger and subsequently to the parser and thus require special treatment. We investigate a specific type of multi-word expressions: binomial adverbs. These adverbs follow the pattern *adverb + conjunction + adverb*. English examples are *by and large*, *first and foremost*, *over and over*. In German we find *ab und zu*, *ganz und gar*, *nach wie vor* (EN: occasionally, completely, still). The most prominent example in Swedish is *till och med*, but there are many others like *blott och bart*, *helt och hållet*, *om och om (igen)* (EN: purely and simply, completely, again and again). We searched manually annotated corpora for English, German and Swedish for occurrences of such binomial adverbs. We also ex-

amined an automatically annotated version of Europarl for these three languages. We found that German and Swedish have more occurrences of binomial adverbs in both the manually annotated corpora and the automatically tagged and parsed Europarl. We will present the comparison across the three languages in sections 2.2 and 2.3.

We will start with a definition of binomial adverbs as a subclass of multi-word adverbs. Section 3 has our results of the parsing evaluation and section 4 describes our machine translation evaluation.

This paper is thus an evaluation paper and not a methodology paper. We identify a subclass of multi-word expressions that need special treatment in order to improve parsing results and machine translation (MT).

2 Multi-word Adverbs

In this paper we focus on a subclass of multi-word adverbs which we call **binomial adverbs** since they intersect with binomials as described by Gereon Müller (1997) and Sandra Mollin (2014). This class of adverbs is interesting since the “adverbs” that make up the construction are often taken from other parts-of-speech. For example, the German binomial adverb *nach und nach* (EN: gradually) is constructed out of conjoined prepositions. The frequent Swedish binomial adverb *till och med* (EN: even, through) is annotated in the Stockholm-Umeå Corpus (SUC) as conjoined adverbs 120 times and as conjoined prepositions 16 times. Certainly the “adverbs” *till* and *med* in this binomial construction are much more frequently used as prepositions and verb particles than as adverbs (see table 2). Sometimes other parts-of-speech are used as e.g. the adjective in English *by and large* or the Swedish noun *hållet* (EN: distance, direction) in *helt och hållet*. These PoS am-

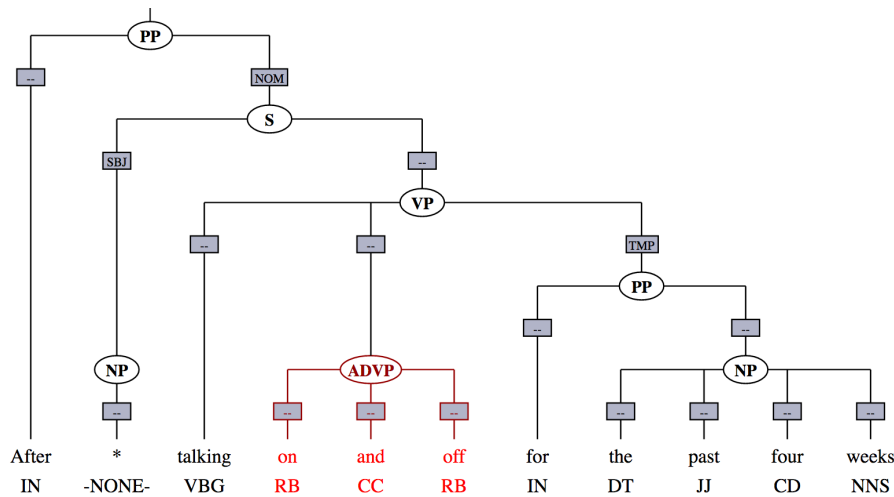


Figure 1: Syntax tree with multi-word adverb (*on and off*) from the Penn treebank. The multi-word adverb is annotated as adverbial phrase (ADVP).

biguities may lead to processing errors in parsing or translation.

2.1 Related work

Binomial constructions have been studied in linguistics for many years. (Bendz, 1965) is an early monograph on “word pairs” in Swedish with comparisons to Danish, English and German. Bendz deals not only with adverbs but all kinds of coordinated word pairs. He presents a semantic classification to distinguish the binomials into

1. opposition pairs (e.g. EN: *sooner or later, to and fro*, DE: *dick und dünn, weit und breit*, SV: *tjockt och tunt, vitt och brett*)
2. enumeration pairs (e.g. SV: *män, kvinnor och barn; tid och rum*), but no examples for adverbs
3. synonym pairs (e.g. EN: *first and foremost, simply and solely*, DE: *frank und frei, ganz und gar*, SV: *blott och bart, helt och hållet*)

Bendz discusses many aspects of word pairs such as inheritance of the constructions from Latin, their prominence in the literature, but also formal properties such as alliteration, assonance and rhyme. He mentions word order constraints and formulates the hypothesis that “the more frequent a word pair becomes, the fixer is word order” (p. 21, translation by the authors). He links the ordering tendencies of unequal words to Behaghel’s law of increasing terms which predicts that shorter words come before longer words. Some examples

from English might illustrate this law: *first and foremost*, **foremost and first*, *far and away*, **away and far*, *now and again*, **again and now*.

Müller (1997) presents a detailed study of binomial constructions in German (e.g. *Fug und Recht, samt und sonders*) which includes binomial adverbs. He is particularly interested in order constraints which he regards as a defining feature of binomial constructions.

Müller elaborates that end rhyme, alliteration (*ganz und gar*) and assonances (the repetition of vowel sounds to create internal rhyming; e.g. *dann und wann*) are typical properties of binomial constructions.

In a recent book Mollin (2014) discusses ordering constraints of English binomials in great detail. Mollin performs a study on the British National Corpus (BNC), a 100 million word corpus collected in the 1990s. She relies on automatic PoS tagging and searches for pairs of the same parts-of-speech like ‘noun *and* noun’, ‘verb *and* verb’, ‘adverb *and* adverb’ etc. She investigated all candidates where the more frequent sequence occurs 50 times or more. This resulted in 544 types. Among these, there are 20 adverbs, with *up and down*, *here and there*, *now and then* being the most frequent ones. Mollin judged the candidates by a so called irreversibility score which is the ratio of the more frequent order against the frequency of both orders. The highest irreversibility scores of 100% go to *back and forth*, *out and about*, *today and tomorrow* all of which were only found in the BNC in the given sequence.

There have been a lot of studies on the automatic parsing and translation of multi-word expressions. But surprisingly few of them deal with multi-word adverbs or binomials. Widdows and Dorow (2005) find idiomatic expressions of the form ‘noun *and/or* noun’ in the BNC by exploiting ordering constraints. They elaborate on the difference between symmetric and asymmetric relationships between nouns, where asymmetric relations (one order being clearly more frequent than the other) may indicate idiomaticity or a number of other constraints such as hierarchies, gender asymmetries or temporal order. Michelbacher et al. (2007) explore the properties of asymmetric association measures which pay tribute to the, often-times, asymmetric nature of collocations.

Volk et al. (2016) have investigated multi-word adverbs for German and their impact on PoS tagging accuracy and the re-combination of separated verb prefixes to their respective verbs. Since some separated verb prefixes are homographs with prepositions (and obviously derived from prepositions) and also used in binomial adverbs (as e.g. *ab und zu*, *nach und nach*), it is important to identify the binomial adverbs in order to avoid confusion with separated verb prefixes and to prevent subsequent erroneous verb lemmas and syntax structures.

Out of the large pool of NLP studies on multi-word expressions in general let us mention Ramisch (2015) who introduces methods for the discovery of multi-word expressions, among others the computation of collocation scores. Constant and Nivre (2016) show how to integrate MWEs into dependency parsing.

Nasr et al. (2015) propose a method for jointly tokenizing and parsing adverb-conjunction combinations in French (*ainsi que*, *bien que*). Their problem is similar to ours in that the combination shows no internal variability which makes it easy to spot but is ambiguous because of cases with literal usage (e.g. *ainsi que* is reported to be a multiword unit in only 76.6% of the cases). Their method relies on subcategorization information for the verbs with respect to the verbs’ tendency to take subordinate *que* clauses. They report on clear improvements in parsing these cases.

From the long history of MWEs in machine translation we refer to two recent works. Bouamor et al. (2012) show how to integrate MWEs into statistical machine translation for French-English.

Tan and Pal (2014) extract MWEs for Hindi and English, integrate them into MT and report on an improvement in MT quality.

2.2 Binomial Adverbs in Manually Curated Corpora

In order to get an overview of the frequency of binomial adverbs we evaluated the Penn Treebank for English, the TIGER treebank for German, and SUC for Swedish (all of which have roughly 1 million tokens).

In the German TIGER treebank we find 211 syntactic constituents labeled as coordinated adverb phrase (CAVP), corresponding to 110 types. The top frequent ones are *nach wie vor* (66 occurrences), *mehr oder weniger* (10), and *nach und nach* (7). Only 26 types occur more than once.

For Swedish we used the Stockholm-Umeå Corpus (SUC) which is a manually checked corpus with lemmas and parts-of-speech. SUC is part of the Swedish treebank. In SUC we find 985 sequences *adverb + conjunction + adverb* which are potential candidates for binomial adverbs. *till och med* is the most frequent sequence with 120 occurrences, followed by *mer eller mindre* (61) and *då och då* (45 occurrences). The prominence of *till och med* is underlined by the fact that it, additionally, occurs in SUC as an acronym with 52 occurrences (spelled as *t o m* or *t.o.m.*). Furthermore *till och med* is in SUC 16 times as a conjunction of prepositions. So, it is truly ambiguous with a clear frequency bias towards being a binomial adverb.

If we broaden the search for “adverb or particle or preposition” in the conjunction pattern, then we get 1510 hits in SUC. An analogous query in the Penn Treebank results in only 238 hits, which gives a first indication that binomial adverbs are more frequent in Swedish than in English. The most frequent hits in the Penn treebank are *up and down* (13), *in and out* (8), and *sooner or later*, *back and forth* (7 each). Given the syntactic annotation we can constrain our search to cases where the sequence has an adverb phrase (ADVP) as mother node which reduces our hits to 115. Figure 1 shows an example tree from the Penn Treebank with the binomial adverb *on and off*.

2.3 Binomial Adverbs in Europarl

We annotated Europarl in order to extract all candidates for binomial adverbs. We took a version of Europarl that has 43.1 million tokens for English, 41.1 million tokens for German and 36.1 million

	EN glosses	EN translation	SUC freq	SUC type
<i>först och främst</i>	first and mainly	first and foremost	12	coord adverb
<i>helt och hållet</i>	whole and distance	completely	12	coord adverb
<i>i och för (sig)</i>	in and for	in itself / actually	36	coord prep
<i>i och med</i>	in and with	since / as a result of	61	coord prep
<i>kors och tvärs</i>	cross and across	criss-cross	6	coord adverb
<i>om och om (igen)</i>	again and again	again and again	9	coord adverb
<i>till och med</i>	to and with	even / until	136	coord adv / prep

Table 1: Idiomatic Swedish binomial adverbs with words that have multiple PoS tags (particles, prepositions, nouns). Frequencies (ignoring case) are from the Stockholm Umeå Corpus (1.16 million tokens, mixed texts).

	preposition PP	adverb AB	particle PL	miscellaneous
<i>för</i>	11,035	401	63	101 KN, 28 SN, 44 VB
<i>i</i>	25,522	-	123	4 misc
<i>med</i>	11,063	166	544	2 misc
<i>om</i>	5011	143	400	1 KN, 2395 SN, 2 misc
<i>till</i>	9500	240	674	

Table 2: Part of Speech tag frequencies in SUC for particles that occur in multi-word adverbs (lower case usage only). Miscellaneous PoS tags include conjunction (KN), subjunction (SN), verb (VB).

tokens for Swedish. We tagged the English and German parts with TreeTagger and parsed with MaltParser. We annotated the Swedish part with a combination of Stagger und Maltparser with the standard model for Swedish¹. All language-specific PoS tags were also mapped to universal PoS tags.

This allows us to extract binomial adverbs for the three languages with the same query. Since we know that adverbs in these constructions are sometimes confused with particles (PRT) and prepositions (ADP), we search for the pattern (ADV or ADP or PRT) followed by a conjunction followed by one of these three PoS tags again (ADV or ADP or PRT). This leads to the following results. For Swedish we find 37,973 occurrences (with 6983 binomial ADV types), while for English we have 23,509 occurrences (with 8034 types). So, given that the Swedish part of our Europarl corpus has 10% less tokens overall than English, this is a clear indication that binomial adverbs are more common in Swedish than in English.

The frequencies for German are not directly comparable with English and Swedish because of

differences in the annotation of adjectives. German adjectives that function as adverbs or predicates are tagged as ADJD (in contrast with attributive adjectives which get the tag ADJA). If we were to map all ADJDs to ADVs in the universal PoS tag set, we would get way more adverbs than in English or Swedish. We therefore decided to skip the ADJD cases and use the same query for German as for the other two languages. This results in 19,427 occurrences (with 3830 binomial ADV types).

A closer look at Swedish binomial adverbs reveals that some of them are tagged as ADJ as well. For example, *helt och hållet* is tagged as “ADV CONJ ADV” 2448 times, but also 98 times erroneously as “ADJ CONJ ADJ”, and 46 times as “ADJ CONJ ADV” (plus 18 times with other miscellaneous tag combinations).

The most frequent ones for Swedish are *till och med*, *först och främst*, *helt och hållet*², while for English they are *whether or not*, *once and for (all)*, *more and more*. The top frequent German candidates are *nach wie vor*, *voll und ganz*, (so) *schnell wie möglich*. The examples show that sometimes

¹http://maltparser.org/mco/swedish_parser/swemalt.html

²The Swedish Europarl has also 1096 occurrences of the acronym *t.o.m.* which stands for *till och med*.

we catch candidates that are part of larger idiomatic expressions (as in the case of *once and for (all)*; *(so) schnell wie möglich* (EN: as fast as possible)).

2.4 Binomial Adverbs and Idiomaticity

The above sections exemplify that many binomial adverbs are true multi-word expressions (with non-compositional semantics) that need special treatment in natural language processing. In order to zoom in on idiomatic binomial adverbs, we used two methods. First we checked for order restrictions. If a candidate “X conjunction Y” has a certain corpus frequency but the opposite order “Y conjunction X” does not occur (or occurs with a much smaller frequency), then this increases the likelihood that the candidate is an idiomatic expression.

This check excludes English candidates like *clearly and fully* (3 times in either order), German candidates like *heute und jetzt* (6 times in either order), and Swedish candidates like *där och när* (4 times in either order), and *alltid och överallt* which occurs 13 times in this order and 7 times in the reverse order in Europarl. In this way we exclude several hundred candidates in each language.

Obviously this method does not work for candidates with word repetitions. Our Europarl search resulted in 47 such reduplication candidates in Swedish³ with *mer och mer*, *då och då*, *om och om* (*igen*), *längre och längre*, *så och så* being the top frequent ones. It is striking that reduplications are often used with comparative forms as e.g. *bättre och bättre*, *snabbare och snabbare*, *mindre och mindre* and with words that already stand for repetitions *åter och åter*, *igen och igen*, *till och till* which are intensified in this way. The observations for English and German with respect to reduplications are very similar.

Second, we computed collocation scores (mutual information scores, MI) for all candidates “X conjunction Y”. For this we used the pair frequencies of “X conjunction” and “conjunction Y” in comparison with the frequency of the triple “X conjunction Y”. Our formula is

$$MI(X, C, Y) = \log_2 \frac{N \cdot f(X, C, Y)}{f(X, C) \cdot f(C, Y)}$$

with C being the conjunction between the particles X and Y, and N being the number of tokens

³Such repetition candidates are called echoics in (Mollin, 2014).

in the corpus. In this way the MI score predicts the probability of “X conjunction” being followed by Y, and the likelihood of “conjunction Y” being preceded by X.

We set a minimum threshold of 6 occurrences for the triple. This cutoff results in 437 English candidate triples. High frequency candidates like *once and for*, *over and above*, *again and again* get high MI scores of 12 to 14. At the lower end we find *here and on*, *up and in* that are certainly not idiomatic multi-word units.

For German high MI examples are *hin oder her* (MI: 21) and *nie und nimmer* (MI: 20.8). Frequent candidates *ganz und gar*, *mehr und mehr*, *nach und nach* receive prominent MI scores between 14 and 16 which increases their likelihood of being multi-word units.

For Swedish the MI scores leave 426 candidate triples in the game. The top frequency candidates *till och med*, *först och främst*, *helt och hållet* get high MI scores of 10 and above, with *kors och tvärs*, *blott och bart*, *sönder och samman* receiving scores of 20+. At the other end of the scale we are able to rule out candidates with scores below 8 such as *här och i*, *nu och då*, *snabbt och på*.

In conclusion, binomial adverbs cover the whole spectrum of idiomaticity and can only be interpreted correctly when their ordering constraints and their collocation strengths are appropriately considered. Following the above considerations we picked 7 candidates per language for evaluation in parsing and machine translation. Criteria for the selection were idiomaticity, frequency and PoS ambiguities. Tables 3, 4 and 5 show the selected candidates.

3 Evaluation of Binomial Adverbs in Parsing

For the evaluation of binomial adverbs we ran MaltParser over the English, German and Swedish parts of the Europarl corpus. We then profiled the parsing results of our selected binomial adverbs. We counted the PoS patterns assigned to these adverbs, their dependency patterns (ignoring the labels) and their dependency label patterns. For example, in figure 2 we find the PoS pattern “ADV - CONJ - ADV” for *on and on* which is the desired tag sequence. Both the conjunction and the second adverb are marked as being dependent on the first adverb. These are the desired dependencies in English and Swedish. The German dependency

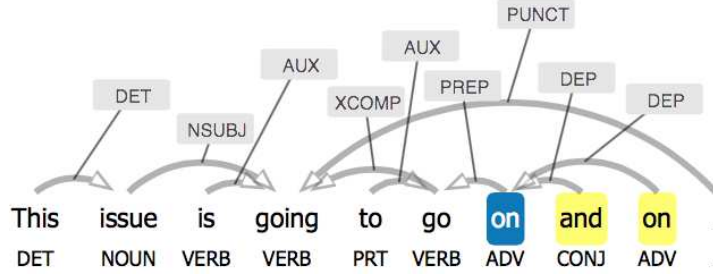


Figure 2: English sentence from Europarl with multi-word adverb (*on and on*) correctly tagged and parsed (disregarding the dependency labels).

	Europarl	PoS acc.	Parsing acc.	MT acc. EN → DE
<i>by and large</i>	161	0.6%	72.1%	100%
<i>first and foremost</i>	1496	17.2%	80.4%	100%
<i>now and again</i>	46	100%	100%	90.0%
<i>on and off</i>	20	55.0%	70.0%	32.0%
<i>on and on</i>	43	46.5%	81.4%	71.4%
<i>out and about</i>	7	85.7%	42.9%	66.7%
<i>over and over</i>	150	52.0%	50.0%	91.7%

Table 3: English binomial adverbs selected for evaluation.

parser marks the conjunction as dependent on the first adverb and the second adverb as dependent on the conjunction. When MaltParser assigns these dependencies, we count the parse of the binomial adverb as correct.

Table 3 shows the results for the English adverbs in our evaluation. For example, *by and large* occurs 161 times in Europarl (counting both upper and lower case occurrences). But only a single occurrence (0.6%) gets the PoS tags “ADV - CONJ - ADV”. Instead 145 occurrences are tagged as “ADP - CONJ - ADJ” (adposition - conjunction - adjective), and 15 occurrences are tagged as “ADP - CONJ - ADV”. Despite the high error rate in PoS tagging, the dependency arcs are correct in 72.1% (disregarding the dependency labels), indicating that this is a multi-word unit. This might be influenced by the fact that we parsed Europarl based on language-specific PoS tags and only later converted them into universal PoS tags.

Tagging the particles as adpositions also accounts for the 53.5% PoS errors with *on and on*. Still we observe a high accuracy of unlabeled dependencies with 81.4%. Figure 2 shows this adverb correctly parsed, while figure 3 has an incorrectly parsed example.

One should note that our evaluation may fall victim to cases of literal, i.e. non-idiomatic, usage.

For example our extracted list includes the sentence ... *which is being worked on and on which we may make progress* But such occurrences of literal usage are rare.

Tables 4 and 5 have the corresponding results for the selected German and Swedish adverbs. We cannot discuss all observations here, but let us focus on the most prominent Swedish binomial adverb.

For *till och med* we expect to get a dependency profile where both *och* and *med* have a “head” dependency to *till*. And *till* then has a dependency as contrastive adverbial (CA), attitude adverbial (MA) or other adverbial (AA) or (seldom) as time adverbial (TA) to the appropriate word in the sentence. The good news is that *till och med* has the desired dependencies in 95.5% of the cases in Europarl.

It is also positive that *till* has plausible dependency labels in about 70% of the cases. However, this leaves 30% of the cases with dubious external dependency labels. For example, there are 4 occurrences of *till och med* with *och* being the root (!) of the sentence. All 4 occurrences are immediately preceded by the conjunction “eller”. There are 162 occurrences where *till* is the root of the sentence which is also unlikely. There are over one thousand occurrences with *till* labeled as time

	EN translation	Europarl	PoS acc.	Parsing acc.	MT acc. DE → EN
<i>ab und an</i>	occasionally	19	0%	100%	80.0%
<i>ab und zu</i>	occasionally	65	0%	98.5%	83.9%
<i>eh und je</i>	ever	24	0%	100%	90.0%
<i>hin und wieder</i>	sometimes	84	98.90%	85.7%	67.7%
<i>kreuz und quer</i>	criss-cross	20	0%	100%	55.0%
<i>nach und nach</i>	gradually	373	85.0%	98.7%	87.1%
<i>nach wie vor</i>	still	4723	99.9%	71.3%	96.8%

Table 4: German binomial adverbs selected for evaluation.

	EN translation	Europarl	PoS acc.	Parsing acc.	MT acc. SV → EN
<i>först och främst</i>	first and foremost	3988	99.4%	92.0%	95.8%
<i>helt och hållet</i>	completely	2610	93.8%	40.0%	100%
<i>i och för (sig)</i>	in itself / actually	249	52.2%	51.4%	48.0%
<i>i och med</i>	since / as / with	3035	67.9%	67.6%	81.8%
<i>kors och tvärs</i>	criss-cross	21	100%	14.3%	33.3%
<i>om och om</i>	again and again	250	100%	54.0%	100%
<i>till och med</i>	even / until	6802	99.9%	95.5%	94.3%

Table 5: Swedish binomial adverbs and conjunction (*i och med*) selected for evaluation.

adverbial which are mostly wrong. So, this calls for a special treatment of the binomial adverbs either prior to parsing or during parsing.

4 Evaluation of Binomial Adverbs in Machine Translation

We also evaluated how well a state-of-the-art MT system, Google Translate, handles the selected binomial adverbs. For each adverb we extracted sentences from various corpora. We did not use Europarl sentences for this evaluation since chances are high that this corpus is part of the training data for Google Translate. We sorted the extracted sentences by length (number of tokens) and dropped the short ones (less than 10 tokens) because they may not provide enough context for MT, and we dropped the long ones (more than 50 tokens) because they may confuse the MT system and because they make manual evaluation more time-consuming. From the remaining sentences we selected 30 per adverb and fed them to Google Translate. We translated EN → DE, DE → EN, and SV → EN and then manually evaluated whether the binomial adverbs were translated correctly. Tables 3, 4 and 5 contain the resulting MT accuracy. The numbers describe the percentages of sentences that had a correct translation of the binomial adverb. It does not mean that the complete sentences were translated correctly.

The first impression is that binomial adverbs are handled surprisingly well by Google Translate. For example, the German *ab und zu* is not only translated correctly but also with some variation into English as *occasional*, *sometimes*, *from time to time* which are all good translations. See example 1 which features a correct translation of the binomial adverb but also a number of tricky word reorderings which result in an excellent rendering of the meaning in English.

- (1) DE: Somit riskieren wir **ab und zu**, im Sande stecken zu bleiben.
 Google EN: Thus we **sometimes** risk getting stuck in the sand.

The most striking problem in Google’s machine translation of the binomial adverbs are omissions. The adverb is sometimes dropped in the translation as in example 2. This is an irritating finding, in particular since the generated target language sentence is fluent and grammatically correct. It looks good at first sight but misses an important aspect (expressed by the multi-word adverb) from the input sentence.

- (2) DE: Ein dritter zeigt **ab und zu** Dias, die er selber in der Umgebung der Hütte gemacht hat.
 Google EN: A third shows slides, which he himself has made in the vicinity of the hut.

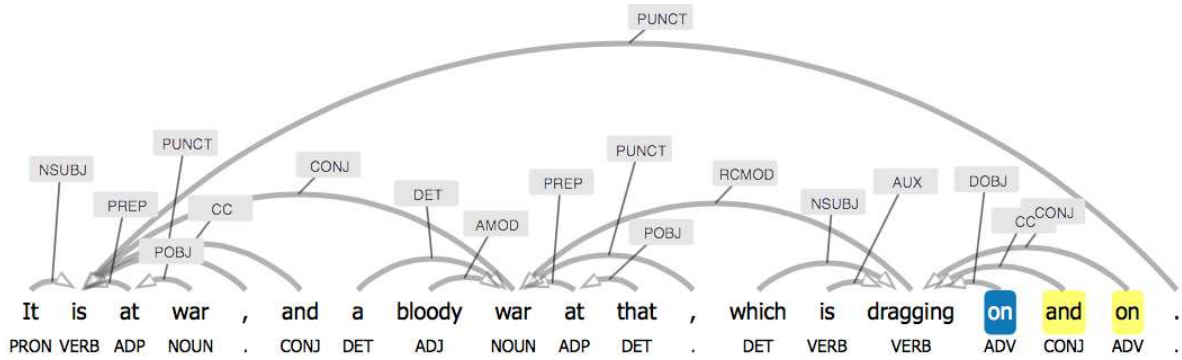


Figure 3: English sentence from Europarl with multi-word adverb (*on and on*) correctly tagged but incorrectly parsed.

Such omissions account for the majority of translation errors with respect to binomial adverbs. There are occasional wrong translations but to a much lesser extent. In example 3 we see that *on and off* is erroneously translated into German. It should have been *ab und zu, mit Unterbrechungen*.

- (3) EN: The rains continued, **on and off** until mid April, unusually late for Jordan.
 Google DE: Der Regen fuhr fort, **ab und ab** bis Mitte April, ungewöhnlich spät für Jordanien.

The translation of the English adverb *on and off* is difficult since sometimes it can have its literal meaning (*a torch flickered on and off*) whereas in other cases only the idiomatic translation is correct. This may explain its low MT accuracy.

For Swedish we checked *till och med* because it is so frequent and it also can serve two purposes. With that in mind we conclude that a 94.3% translation accuracy is good. In addition, we randomly extracted 10 sentences where *till och med* is an adverb and 10 sentences where the sequence is a conjoined preposition. Interestingly, the MT system translated *till och med* correctly in all 20 test sentences. The ones with the adverb reading were all translated with the English word *even* whereas the preposition cases were translated with *to / until / through*.

5 Conclusion

We have shown how to narrow down the search for binomial adverbs, a special type of multi-word expressions. We used the irreversibility score and a mutual information score to find cases that are top candidates for idiomatic usage.

We subsequently selected 7 such binomial adverbs from English, German and Swedish each and evaluated them in PoS tagging, dependency parsing and machine translation. The results are mixed in that PoS tagging and parsing works very good for some and badly for others. If we consider that down-stream applications rely on the parsing results, our study pin-points the need to handle such binomial adverbs with more care.

Statistical and Neural Machine Translation do not rely on parsing, and we therefore evaluated the binomial adverbs separately with Google Translate. We observed that frequent binomial adverbs like *by and large*, *first and foremost*, *over and over* in English, or *nach und nach*, *nach wie vor* in German, or *helt och hållet*, *till och med* in Swedish are translated well but not perfectly. The biggest problem is that Google Translate sometimes omits the binomial adverb which can be detrimental for the understanding of the sentence in the target language.

There is currently no repository of English, German or Swedish multi-word adverbs as in French (Laporte and Voyatzi, 2008) and some other languages. Our work would like to contribute to compiling such repositories.

Acknowledgments

This research was supported by the Swiss National Science Foundation under grant 105215_146781 for “SPARCLING: Large Scale PARallel Corpora for LINGuistic Investigation” (2013-2017). Thanks to Christof Bless for his tool to visualize the dependency trees. Thanks also to the anonymous reviewers for helpful comments on an earlier version of this paper.

References

- Gerhard Bendz. 1965. *Ordpar*. P. A. Nordstedt & Söners Förlag, Stockholm.
- Dhouha Bouamor, Nasredine Semmar, and Pierre Zweigenbaum. 2012. Identifying bilingual multiword expressions for statistical machine translation. In *Proc. of LREC*. Istanbul.
- Mathieu Constant and Joakim Nivre. 2016. A transition-based system for joint lexical and syntactic analysis. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. Berlin.
- Eric Laporte and Stavroula Voyatzi. 2008. An electronic dictionary of French multiword adverbs. In *Proc. of LREC*. Marrakech, Morocco.
- Lukas Michelbacher, Stefan Evert, and Hinrich Schütze. 2007. Asymmetric association measures. In *Proceedings of Conference on Recent Advances in Natural Language Processing (RANLP)*.
- Sandra Mollin. 2014. *The (Ir)reversibility of English Binomials. Corpus, constraints, developments*, volume 64 of *Studies in Corpus Linguistics*. John Benjamins.
- Gereon Müller. 1997. Beschränkungen für Binomialbildungen im Deutschen. *Zeitschrift für Sprachwissenschaft* 16(1):25–51.
- Alexis Nasr, Carlos Ramisch, José Deulofeu, and André Valli. 2015. Joint dependency parsing and multiword expression tokenization. In *Proceedings of ACL*. Beijing, pages 1116–1126.
- Carlos Ramisch. 2015. *Multiword Expressions Acquisition: A Generic and Open Framework*. Theory and Applications of Natural Language Processing. Springer.
- Liling Tan and Santanu Pal. 2014. Manawi: Using multi-word expressions and named entities to improve machine translation. In *Proceedings of the Ninth Workshop on Statistical Machine Translation*. Baltimore.
- Martin Volk, Simon Clematide, Johannes Graën, and Phillip Ströbel. 2016. Bi-particle adverbs, PoS-tagging and the recognition of German separable prefix verbs. In *Proceedings of KONVENS*. Bochum.
- Dominic Widdows and Beate Dorow. 2005. Automatic extraction of idioms using graph analysis and asymmetric lexicosyntactic patterns. In *Proceedings of the ACL-SIGLEX 2005 Workshop on Deep Lexical Acquisition*. Ann Arbor, Michigan.